

Metadata

Dataset Title

LAGOS-US Race and Ethnicity: Dataset classifying conterminous US lake communities based on racial and ethnic demographic data (v1.0)

Short name or nickname you use to refer to this dataset:

LAGOS-RE

Abstract

Knowing the consistency of water quality sampling in lakes surrounded by a variety of racial and ethnic communities is important when thinking about potential policy uses and community impacts. By using the 2010 US Census race and ethnicity demographic tract data, we analyzed the frequency (i.e., number of years and consistency) of lake water quality sampling according to racial and ethnic demographics in surrounding neighborhoods. Our approach classified human communities near lakes as predominantly White or people of color (POC), and Hispanic or non-Hispanic. Associated R analysis scripts can also be found in this folder. Our data and approach can be used for future studies seeking to analyze environmental monitoring practices in relation to human demographic variables, particularly from US Census data.

Investigators

First Name	Middle Initial	Last Name	Organization	e-mail address	ORCID ID (optional)
Jessica		Díaz Vázquez	Michigan State University	jessica.dv405@gmail.com	0000-0001-8493-4035
Ian	M	McCullough	Michigan State University	immccull@gmail.com	0000-0002-6832-674X
Maggie		Haite	Michigan State University	haitemar@msu.edu	0000-0002-7490-8850
Patricia	A	Soranno	Michigan State University	soranno@msu.edu	0000-0003-1668-9271
Kendra		Spence Cheruvelil	Michigan State University	ksc@msu.edu	0000-0003-1880-2880

Other personnel names and roles

First Name	Middle Initial	Last Name	Organization	e-mail address	ORCID ID (optional)	Role in project
Austin		Delaney	University of Wisconsin	addelany@wisc.edu		LAGOS-US-LIM NO data module curator

License

[CCBY – requires attribution](#)

Keywords

race, ethnicity, environmental justice, lakes, freshwater, limnology, conterminous US, LAGOS

Funding of this work:

Add rows to table if several grants were involved, list only the main PI, start with main grant first:

PI First Name	PI Middle Initial	PI Last Name	PI ORCID ID (optional)	Title of Grant	Funding Agency	Funding Identification Number
Kendra		Spence Cheruvellil	0000-0003-1880-2880	A macrosystems ecology framework for continental-scale prediction and understanding of lakes	US National Science Foundation	US NSF Macrosystems Biology Program grants: DEB-1638679
Jessica		Diaz	0000-0001-8493-4035	College of Agriculture and Natural Resources Undergraduate Research Program	Michigan State University	none
Maggie		Haite	0000-0002-7490-8850	Honors College Professional Assistantship Program	Michigan State University	none

Timeframe

- Begin date: 1970
- End date: 2016
- Data collection ongoing/completed: completed

Geographic location

- Verbal description: conterminous US (lower 48 states and the District of Columbia)
- North bounding coordinate (decimal degree): 49
- South bounding coordinate (decimal degree): 25
- East bounding coordinate (decimal degree): -125
- West bounding coordinate (decimal degree): -67

Taxonomic species or groups

N/A

Methods

[lagoslakeid](#) was obtained from merging the LAGOS-US LIMNO and LAGOS-NE v1.087.3 databases.

[lake_centroidstate](#) is the two-letter postal abbreviation of the state containing the lake central point and

was obtained from LAGOS-US LOCUS v 1.0.

[lake_namelagos](#) was obtained from LAGOS-US LOCUS v 1.0. and is the lake name from a combination of data sources; examples are GNIS, WQP, etc.

[lake_region](#) was obtained based on state boundaries, US Census subregions, and distribution of lakes. The US Census Bureau recognizes 4 regions (Northeast, Midwest, South, and West) and 9 divisions or subregions (New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain and Pacific). In some instances, we kept entire regions and in others we combined subregions. We designated each lake with a region based on the following:

[lake_centroidstate](#) identifier: The six regions contained 14,000 - 33,000 lakes and were:

1. Northeast: Connecticut (CT), Maine (ME), Massachusetts (MA), New Hampshire (NH), New Jersey (NJ), New York (NY), Pennsylvania (PA), Rhode Island (RI), Vermont (VT)
2. Midwest East: Illinois (IL), Indiana (IN), Michigan (MI), Ohio (OH), Wisconsin (WI)
3. Midwest West: Iowa (IA), Kansas (KS), Minnesota (MN), Missouri (MO), Nebraska (NE), North Dakota (ND), South Dakota (SD)
4. Southeast: Alabama (AL), Delaware (DE), Florida (FL), Georgia (GA), Kentucky (KY), Maryland (MD), Mississippi (MS), North Carolina (NC), South Carolina (SC), Tennessee (TN), Virginia (VA), West Virginia (WV), and District of Columbia (DC)
5. South Central: Arkansas (AR), Louisiana (LA), Oklahoma (OK), Texas (TX)
6. West: Arizona (AZ), California (CA), Colorado (CO), Idaho (ID), Montana (MT), New Mexico (NM), Nevada (NV), Oregon (OR), Utah (UT), Washington (WA), Wyoming (WY)

[lake_race](#) and [lake_ethnicity](#) were calculated based on the 2010 US Census race and ethnicity tract-level population data obtained from the National Historical Geographic Information System (NHGIS). Tracts are a subunit of counties and contain from 1,500 - 8,000 people. We used tracts as they are relatively homogeneous in terms of population size across the US, yet are at a higher spatial resolution than the county level. We chose not to use the smallest geographical unit, blocks or block groups, as they are delineated based on geographical features (i.e., roads) rather than population. First, in R, a tract was designated as "POC" (people of color) if the combined population of Black/African American, American Indian/Alaskan Native, Asian, and Native Hawaiian/Pacific Islander was $\geq 25\%$; all other tracts not designated "POC" were defined as "White". A tract was designated as "Hispanic" if the Hispanic/Latino population was $\geq 25\%$; all other tracts not designated as "Hispanic" were defined as "non-Hispanic". Then we used ArcMap 10.6 to designate each lake. We used the Join Attribute Data in tract polygon feature to join the tract-designated race & ethnicity data (previously modified in R). Then, converted the tract polygons to raster data using the Polygon to Raster tool. The input feature was the race and ethnicity tract polygon feature, the value field was "race" when running the process for the race designations and "ethnicity" for the ethnicity designations. We used 90 m as the minimum cell size to ensure that the smallest tract filled at least one full pixel. The 90 m value was calculated by taking the square root of the lowest value in the shape_area field of the joined tract polygon feature. We then applied the Zonal Statistics as a Table tool using the 500 m lake buffer 4 ha polygons from LAGOS US

LOCUS v 1.0 (Smith et al. 2021) as the input raster or feature data, [lagoslakeid](#) identifier as the zone field, the rasterized race or ethnicity data as the input value raster, and MAJORITY as the statistics type.

[date_sampled](#), [year_sampled](#), [chla_sampled](#), [secchi_sampled](#), and [tp_sampled](#) were derived from two databases from the LAke multi-scaled GeOSpatial and temporal database (LAGOS) project (<https://lagoslakes.org/>). The LAGOS-NE LIMNO v1.087.3 database includes water quality data from 87 state, federal, citizen science, non-profit, tribal and university monitoring programs for lakes ≥ 4 hectares across 17 states in the Northeastern and Midwestern US (Soranno et al., 2019). For the remaining 31 US conterminous states and the District of Columbia, we used a preliminary version of LAGOS US LIMNO, which contains water quality data downloaded from the Water Quality Portal (<https://www.waterqualitydata.us>). We selected occurrences where [date_sampled](#) was between June 15 and September 15 (generally coinciding with lake stratification and the peak period of data collection) over the years 1970-2016. Then, we created [year_sampled](#) by selecting only the year from [date_sampled](#). For [chla_sampled](#), [secchi_sampled](#), and [tp_sampled](#), we assigned “YES” if there was a corresponding water quality sample (i.e., chlorophyll-*a*, Secchi disk depth, total phosphorus, respectively) in the preliminary LAGOS US LIMNO data for that [date_sampled](#) and “NO” for all other occurrences.

Data Table

Table name: lake_raceethnicity

Table description: Race and ethnicity designations for lakes ≥ 4 ha within the conterminous US. Includes lake identifiers and data on presence or absence of water quality measures of chlorophyll-*a*, Secchi disk depth (water clarity), and total phosphorus for every lake as well as the date and year sampling occurred.

Column name	Description	Unit or code explanation or date format	Empty value code
lagoslakeid	unique lake identifier developed by LAGOS US LOCUS v1.0 and LAGOS-NE	no units	no empty values
lake_centroidstate	two-letter postal abbreviation of the state containing the lake centroid; from LAGOS US LOCUS v1.0	no units	no empty values
lake_namelagos	lake name from a combination of data sources; examples are GNIS, WQP, etc.; from LAGOS US LOCUS v1.0	no units	NULL
lake_region	the numerical abbreviation of the region containing the lake	1=Northeast, 2=Midwest East, 3=Midwest West, 4=Southeast, 5=South Central,	no empty values

		6=West	
lake_race	the classification of a lake into either POC or White	1 = White, 2 = POC	NULL
lake_ethnicity	the classification of a lake into either H or NH	1 = Non-Hispanic, 2 = Hispanic	NULL
date_sampled	the date sampling occurred for the corresponding water quality sample	yyyy/mm/dd	NULL
year_sampled	the year sampling occurred for the corresponding water quality sample	yyyy	NULL
chla_sampled	an in situ chlorophyll- <i>a</i> sample was collected on this date	Y = yes, N = no	no empty values
secchi_sampled	an in situ Secchi disk depth (water clarity) sample was collected on this date	Y = yes, N = no	no empty values
tp_sampled	an in situ total phosphorus sample was collected on this date	Y = yes, N = no	no empty values

Articles

Article DOI or URL (DOI is preferred)	Article title	Journal title
TBD - upon acceptance/publication	Freshwater ecosystems are monitored disproportionately less in communities of color	tbd

Scripts/code (software)

File name	Description	Scripting language
LAGOS_RE_CODE.R	Data manipulation code to classify lakes ≥ 4 ha in the conterminous US by surrounding population demographics. Connects water quality data with each lake.	R

Data provenance

Dataset title	Dataset DOI or URL	Creator (name & email)	Contact (name & email)
Nhgis0013_ds172_2010_tract (2010 tract Race and Ethnicity)	http://doi.org/10.18128/D050.V16.0	IPUMS NHGIS: Steven Manson, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles.	nhgis@umn.edu

LAGOS-US LOCUS v1.0	https://portal.edirepository.org/nis/mapbrowse?packageid=edi.854.1	Smith, N.J., K.E. Webster, L. Rodriguez, K.S. Cheruvilil, and P.A. Soranno.	ksc@msu.edu
LAGOS-NE _{limno}	https://doi.org/10.6073/pasta/08c6f9311929f4874b01bcc64eb3b2d7	Soranno, P.A., N.R. Lottig, A.D. Delany, and K.S. Cheruvilil.	ksc@msu.edu

Notes and Comments

References

1. Cheruvilil, K.S., Soranno, P.A., McCullough, I.M., Webster, K.E., Rodriguez, L.K. and Smith, N.J. (2021), LAGOS-US LOCUS v1.0: Data module of location, identifiers, and physical characteristics of lakes and their watersheds in the conterminous U.S.. *Limnol Oceanogr Lett*, 6: 270-292.
<https://doi.org/10.1002/lol2.10203>
2. Smith, N.J., K.E. Webster, L.K. Rodriguez, K.S. Cheruvilil, and P.A. Soranno. 2021. LAGOS-US LOCUS v1.0: Data module of location, identifiers, and physical characteristics of lakes and their watersheds in the conterminous U.S. ver 1. EDI.
<https://doi.org/10.6073/pasta/e5c2fb8d77467d3f03de4667ac2173ca>
3. Soranno, P.A., et al. 2017. LAGOS-NE: a multi-scaled geospatial and temporal database of lake ecological context and water quality for thousands of US lakes. *GigaScience* 6, gix101.
<https://doi.org/10.1093/gigascience/gix101>